NPS55Bn74061

NAVAL POSTGRADUATE SCHOOL

Monterey, California

TWO SEQUENTIAL CEP TESTS

by

D. R. Barr

June 1974

Approved for public release; distribution unlimited

Prepared for:

N... l Electronic System Command
...ngton, D. C. 20360

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Rear Admiral Mason Freeman
Superintendent

Jack R. Borsting
Provost

The work reported herein was supported in part by the Office of Naval Electronic System Command, Washington, D. C. 20360.

This report was prepared by:

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NPS55Bn74061	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) TWO SEQUENTIAL CEP TESTS		5. TYPE OF REPORT & PERIOD COVERED Interim; second half of FY74
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) D. R. BARR		8. CONTRACT OR GRANT NUMBER(s) "Statistical Modeling for Bombing Accuracy Evaluation" WR4-6138
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93940		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS WR4-6138
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Electronic Systems Command Washington, D. C. 20360		12. REPORT DATE
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Circular error probable
Sequential testing
Nonparametric Test
Rayleigh distribution

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Two sequential procedures for testing hypotheses about the circular error probable (CEP) of a weapon system under fixed conditions are discussed. One of the tests is the sequential probability ratio test based on a Rayleigh model for radial miss distances. The second procedure is nonparametric, and is essentially a "control chart" approach based upon the proportions of weapon impacts observed inside the specified CEP. Average sample sizes required for the procedures are compared as well as operating characteristics.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601 |

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

I. INTRODUCTION.

Testing hardware systems to determine whether they satisfy operational specifications is frequently an expensive and time-consuming task. In some circumstances, it is possible to decrease the consumption of resources in testing by using sequential test procedures. A sequential test requires tnat limits be placed upon the desired precision of the test (probabilities of errors), and that the alternatives under test (hypotheses) be clearly specified. Such a test with two hypotheses is conducted by taking one observation (test value) at a time and, after each observation, making one of the decisions

(a) terminate and reject the null hypothesis;

(b) terminate and reject the alternate hypothesis;

(c) take another observation.

Since a sequential procedure requires that a decision (a - c above) be made after each observation, it is necessary to be able to obtain the observation value before another test is run (assuming it does not terminate at that point). In the case of testing a bombing system where several test runs may be scheduled in a relatively short span of time, this means that the measure being used in the test must essentially be obtainable in real time, without appreciable processing delays. Similarly, some "field computation", again without long delays, may be required in order to reach one of the decisions a - c.

In what follows we discuss two sequential test procedures for testing whether a specified CEP (or less) has been achieved by a weapon system under given test conditions. The alternate hypothesis is that the CEP is a second, larger, value or more. These values are obtained from contract specifications and operational considerations. One of the procedures is based on the assumption that bomb impacts follow a circular normal distribution, so that radial miss distances are Rayleigh distributed. Even though this parametric model is frequently used in practice, its adequacy depends upon how well the Rayleigh fits the real situation. There is mounting evidence [1] that the Rayleigh model does not fit actual drop data in the upper tail of the distribution. Hence estimates (such as $\widehat{CEP} = 1.1774\ \hat{\sigma}$) and tests based on this model may not perform well, or even acceptably. In addition, the amount of processing required to obtain each miss distance may involve an intolerable delay before a decision is reached about taking another observation. For these reasons, we give below, only a mathematical sketch of the corresponding sequential test, which we call the "Rayleigh sequential test."

We describe the second test procedure as "nonparametric," in that it essentially does not depend on assumptions about the distribution of bomb impact points about the target. This procedure utilizes the observed proportion of impacts that fall within the specified CEP. It is assumed that judgement about whether a given bomb fell within the specified CEP can be made without significant processing delay, so that it will be known shortly after a given bomb impact is observed, whether another test run is required. This sequential procedure is mathematically equivalent to a well known procedure frequently used in reliability testing and quality assurance applications; in what follows we call it the "sample proportion test."

II. DESCRIPTION OF TESTS.

We use the Sequential Probability Ratio Test (SPRT) developed by Wald [2]. The test may be described in terms of a random walk π_n and bounds $B < A$, such that the test continues so long as π_n remains between B and A. The statistic π_n is based on the likelihood ratio, which in turn depends upon the statistical model (in our case, Rayleigh or binomial) and the CEP values specified under the hypotheses. We shall employ the usual approximations to the stopping bounds B and A,

$$B = \frac{\beta}{1-\alpha} \; ; \quad A = \frac{1-\beta}{\alpha} \; ,$$

where α and β are the type I and type II error rates, respectively. There is strong evidence [3] that these approximations are quite adequate, except possibly for very small α and β.

A. Rayleigh Test. It is convenient to state the test in terms of CEP^2 and squared radial miss distances, R^2. Since the original quantities are nonnegative, this test is equivalent to what would be obtained with CEP and R_i. Now assume the Rayleigh model for R, so that

$$R^2 \sim F_{R^2}(y) = 1-e^{-y/2\sigma^2} \quad ; \quad y > 0 \; , \tag{1}$$

an exponential distribution with parameter $\lambda = 1/2\sigma^2$. Hypothesizing $CEP = C$ is equivalent to setting $\lambda = \alpha/C^2$ (say).

Let C_0 denote the squared CEP specified and C_1 denote the squared CEP alternative, where (C_0,C_1) is an indifference zone. Presumably we are indifferent to which hypothesis is rejected when in fact the true

CEP squared is between C_0 and C_1 . In order to test $CEP^2 \le C_0$ against $CEP^2 \ge C_1$, we test the simple hypotheses

$$H_0 : CEP^2 = C_0 \quad \text{vs} \quad H_1 : CEP^2 = C_1 \tag{2}$$

The SPRT continues so long as

$$B < \prod_{i=1}^{n} \frac{f(y_i \; ; C_1)}{f(y_i \; ; C_0)} < A , \tag{3}$$

where $f(y_i \; ; C_j)$, the density corresponding to (1) with $\frac{1}{2\sigma^2} = \alpha/C_j$, is

$$f(y_i \; ; C_j) = \frac{\alpha}{C_j} \exp(- \alpha_i/C_j) . \tag{4}$$

Taking logs in expression (3), with the explicit expression (4) for f, the test may be described as follows:

<u>SEQUENTIAL RAYLEIGH TEST</u>: In order to sequentially test hypotheses (2) with approximate size α and power $1-\beta$ when sampling from density (4), continue taking observations R^2 so long as their sum (the sum of the first n observations) falls between the acceptance numbers a_n and the rejection numbers r_n, where

$$a_n = (n \ln\frac{C_1}{C_0} + \ln\frac{\beta}{1-\alpha})\Big/[\frac{1}{C_0} - \frac{1}{C_1}] \; \alpha \quad ,$$

$$r_n = (n \ln\frac{C_1}{C_0} + \ln\frac{1-\beta}{\alpha})\Big/[\frac{1}{C_0} - \frac{1}{C_1}] \; \alpha \quad .$$

Terminate and accept H_0 whenever $\sum_{r=1}^{n} R_r^2 \le a_n$; terminate and reject H_0 whenever $\sum_{r=1}^{n} R_r^2 \ge r_n$.

The operating characteristic L of this test is approximately

$$L(C) = P[\text{Test accepts } H_0 \mid CEP^2 = C] \approx (A^{h(C)} - 1)/(A^{h(C)} - B^{h(C)}) ,$$

where $h(C)$ is any (positive) solution to the equation

$$C = \frac{\left(\frac{C_0}{C_1}\right)^{h(C)} - 1}{h(C)\left[\frac{1}{C_1} - \frac{1}{C_0}\right]}$$

Let N denote the (random) sample size at termination. The average sample size required for termination when $CEP^2 = C$ is approximately

$$E(N \mid CEP^2 = C) \approx \frac{[1-L(C)]\ln\frac{1-\beta}{\alpha} + L(C)\ln\frac{\beta}{1-\alpha}}{\ln\frac{C_o}{C_1} + C\left[\frac{1}{C_0} - \frac{1}{C_1}\right]}$$

B. Sample Proportion Test. In order to test the hypotheses (2) without making strong assumptions about the distribution of impact points about the target, one may transform each data point by an indicator function which gives "hit-miss" data. Suppose, then, that X_i is 1 if $R_i \leq \sqrt{C_0}$; $X_i = 0$ otherwise. Assuming the miss distances $R_1, R_2, \ldots$ are a random sample from an (unknown) continuous distribution, $X_1, X_2, \ldots$ are IID Bernoulli random variables under H_0 and H_a . For these transformed variables the null hypothesis $H_0 : CEP^2 \leq C_0$ is equivalent to $H_0' : p \geq 1/2$, where p is the "success" parameter, $p = P[X = 1]$. We also need to determine limits on p under $H_a : CEP^2 \geq C_1$. This requires a model which relates $P[R_i \le \sqrt{C_0}\,]$ to C_1, through the distribution of R_i. For this purpose, we assume a Rayleigh formula. However, this is not a strong assumption here (in contrast to the parametric model discussed above), because we do not rely on the formula for values far from the hypothesized medians of the distributions. It has been noted above that the Rayleigh model gives adequate fit for these values; it is the <u>tail</u> behavior of the Rayleigh that seems to invalidate its use for many CEP estimation and testing problems. Since we are using the Rayleigh formula for moderate values, and in fact using it only to assess $P[X_i = 1 \mid CEP^2 = C_1]$, we feel this is not a serious assumption and that it is not misleading to refer to the present test as a "nonparametric" test. Under this assumption and with the transformed variables X_i, the alternate hypothesis $H_a : CEP^2 \ge C_1$ is equivalent to $H_a^{'} : p \le p_1$, where

$$p_1 = P[X_i = 1 \mid CEP^2 = C_1] = P[R_i^2 \le C_0 \mid CEP^2 = C_1]$$

$$= 1 - \exp\left(-C_0 \quad \alpha\right)$$

$$= 1 - 2^{-C_0/C_1}.$$

The likelihood ratio in this case is

$$\prod_{i=1}^{n} \frac{f(X_i\ ;\ p_1)}{f(X_i\ ;\ 1/2)} = \left(\frac{p_1}{1-p_1}\right)^{\sum_{i=1}^{n} X_i} \cdot [2(1-p_1)]^n ,$$

which remains between B and A so long as $\sum_{i=1}^{n} X_i$ is between

$$a_n = \left(\ln \frac{\beta}{1-\alpha} + n\left(\frac{C_0}{C_1} - 1\right) \alpha \right) / \ln\left(2^{C_0/C_1} - 1\right) \tag{4}$$

and

$$r_n = \left(\ln \frac{1-\beta}{\alpha} + n \left(\frac{C_0}{C_1} - 1 \right) \alpha \right) / \ln\left(2^{C_0/C_1} - 1 \right),$$

the "acceptance number" and "rejection number," respectively. Note that the test is dependent upon C_0 and C_1 only through their ratio, C_0/C_1.

We can now state our "nonparametric" sequential CEP test as follows:

<u>SAMPLE PROPORTION TEST</u>. In order to sequentially test hypotheses (2) with approximate size α and power $1-\beta$, make independent observations of whether each bomb radial misdistance is within $\sqrt{C_0}$ (so $X_i = 1$) or is not (so $X_i = 0$). Continue taking observations X_i so long as $a_n < \sum_{i=1}^{n} X_i < r_n$ where a_n and r_n are given in (4). Terminate and accept H_0 whenever $\sum_{i=1}^{n} X_1 \leq a_n$; terminate and reject H_0 whenever $\sum_{i=1}^{n} X_i \geq r_n$.

The operating characteristic L of this test is approximately

$$L(C) = P(\text{test accepts } H_0 \mid CEP^2 = C) \approx (A^{h(C)} - 1)/(A^{h(C)} - B^{h(C)}) ,$$

where $h(C)$ is any positive solution to the equation

$$C = - C_0 \alpha / \ln(1-J) ,$$

where

$$J = \frac{1 - 2^{h(C)}\left(2^{-C_0/C_1}\right)^{h(C)}}{2^{h(C)}\left(1-2^{-C_0/C_1}\right)^{h(C)} - 2^{h(C)}\left(2^{-C_0/C_1}\right)^{h(C)}}$$

The expected sample size N required for termination of this test is approximately

$$E(N \mid CEP^2 = C) = \frac{[1-L(C)]\ln\frac{1-\beta}{\alpha} + L(C)\ln\frac{\beta}{1-\alpha}}{\left(1-2^{-C_0/C}\right)\ln\left(2-2^{1-C_0/C_1}\right) + 2^{-C_0/C}\alpha^{1-C_0/C_1}}$$

The exact operating characteristic and mean sample size functions do not differ substantially from the above approximations when α and β are not below .1 [3]. The variance in sample size, $V(N \mid CEP^2 = C)$ apparently does not admit such good, simple approximations. However there is empirical evidence [3] that the standard deviation of N tends to be of the order of magnitude of the mean of N. This means that, occasionally, sampling may continue substantially beyond $E(N)$. On the other hand, it is known [4] that the SPRT achieves the smallest average sample size of any test of these hypotheses operating with the same error rates α and β.

III. PERFORMING THE TESTS.

We remarked above that, for the proportion test, the acceptance number and rejection number depend on C_0 and C_1 only through the ratio C_0/C_1. This implies that performance of the proportion test, as well as its operating characteristic and average sample size functions, depend upon the hypothesized values only through C_0/C_1. A similar remark can be made for the Rayleigh test inasmuch as CEP^2 may be viewed as a scale parameter for the distribution of squared radial miss distance R^2. This means that one can, in effect, measure the squared miss distances R_i^2, and the hypothesized values C_0 and C_1 in "C_0-units". In such a case, measurements R_i^2 are converted to $T_i^2 = R_i^2/C_0$, C_0 is taken to be $C_0^1 = C_0/C_0 = 1$, and C_1 is converted to $C_1^1 = C_1/C_0$. Such a point of view allows one to use tabulations of a number of "standard" tests in order to perform tests of a much wider variety of hypotheses.

To illustrate this point, we consider one example of each of the tests described above. For the case $C_1/C_0 = 2$ and $\alpha = \beta = .1$, we present graphical aids which would simplify performance of the tests, as well as graphs of the operating characteristic curves and expected sample size curves. For the Rayleigh case, we also show a sample size variance curve obtained through simulation.

A. Example Sequential Rayleigh Test. Suppose we wish to test $H_0 : CEP^2 \leq 1000$ vs. $H_a : CEP^2 \geq 2000$ with $\alpha = \beta = .10$. With $C_1/C_0 = 2$, we may take $C_0 = 1$ and $C_1 = 2$ as discussed above, with the agreement that the squared radial miss distances actually observed are transformed by dividing by 1000 before computing the test statistic

$\sum_{i=1}^{n} R_i^2$. Equivalently, (and more simply), we can compute the test statistic with the "raw" data actually observed and transform the acceptance and rejection numbers by multiplying by 1000. Thus the test continues so long as ΣR_i^2 is between

$$a_n = 1000 \,[n\alpha - \ln 9] \,/\, \tfrac{1}{2}\alpha = 2000n - 6339.85$$

$$r_n = 1000 \,[n\alpha + \ln 9] \,/\, \tfrac{1}{2}\alpha = 2000n + 6339.85$$

Plotted as functions of n , these bounds are parallel lines (as shown below) having slope 2000 and intercepts $\mp$ 6339.85.

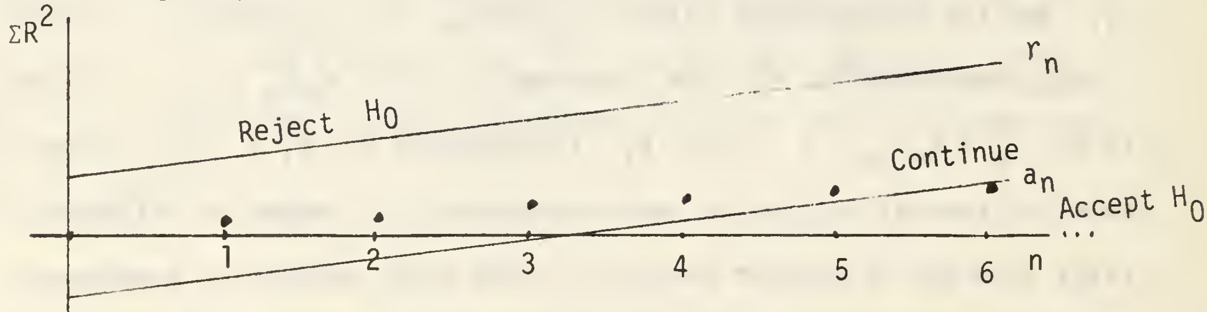

FIGURE 1

A hypothetical sequence of observed values of ΣR_i^2 is shown plotted against n (shown as dots in the figure); that is, the points $(1, y_1)$, $(2, y_1 + y_2)$,..., $(6, \sum_1^6 y_i)$. In performing the test, these points would have been sequentially plotted on a previously prepared "control chart" with the parallel a_n and r_n lines. Note that the hypothetical test terminated with the 6th observation, since the walk $(n, \Sigma R_i^2)$ left the continuation region for the first time at $n = 6$; the terminal decision was to accept H_0 with the example shown. The operating characteristic curve, expected sample size curve and variance in sample size curve for this test are shown in figure 2, as functions of C/C_0 .

Note that the operating characteristic curve passes through (or nearly so) the points $(\alpha, L(1))$ and $(1-\beta, L(2))$, as should be the case.* Note also that mean sample size is largest for C/C_0 values between 1 and 2, again as should be the case, because such values would typically give samples which are not more compatible with one of the hypotheses than the other so no early decision is reached. Similar observation can be made with respect to variance in sample size.

B. Example Sequential Proportion Test. The sequential test based upon "hit - miss" data can also be conveniently performed with a control chart. With the values of (α, β) C_0 and C_1 proposed in the preceding example, we have for this test the hypotheses $H_0 : p \geq .5000$ vs. $H_a : p \leq p_1 = 1 - 1 / \sqrt{2} = .2929$. The acceptance and rejection numbers in this case are

$$a_n = [-\ln 9 - n\alpha/2] / \ln[\sqrt{2} - 1] = .3932n + 2.4929$$

$$r_n = [\ln 9 - n\alpha/2] / \ln[\sqrt{2} - 1] = .3932n - 2.4929$$

which as functions of n are lines with slope .3932 and intercepts $\pm$ 2.4929. In this case the control chart is as shown in figure 3.

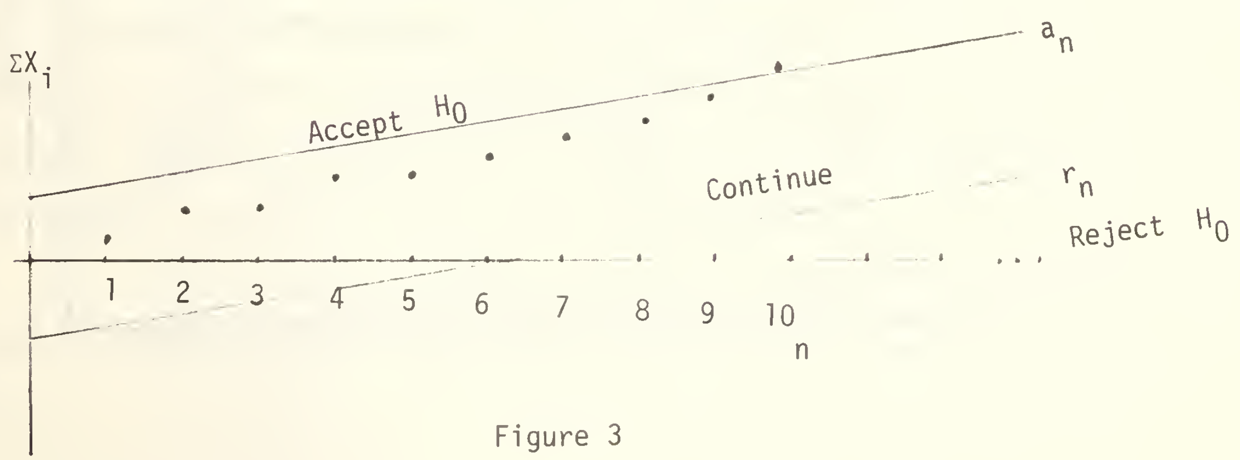

Figure 3

*The OC curve does not pass precisely through these points because of the Wald approximation of B and A. These curves are based on empirical values determined through simulation by Gavlak [3].

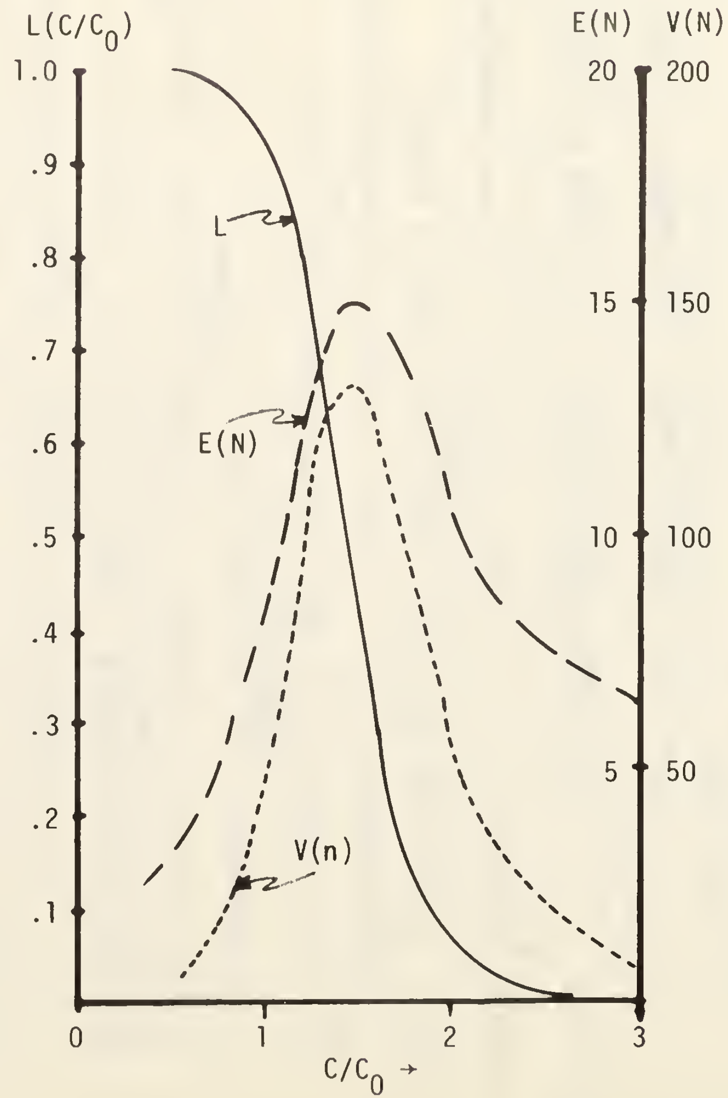

Figure 2. Characteristics of Example Sequential Rayleigh Test.

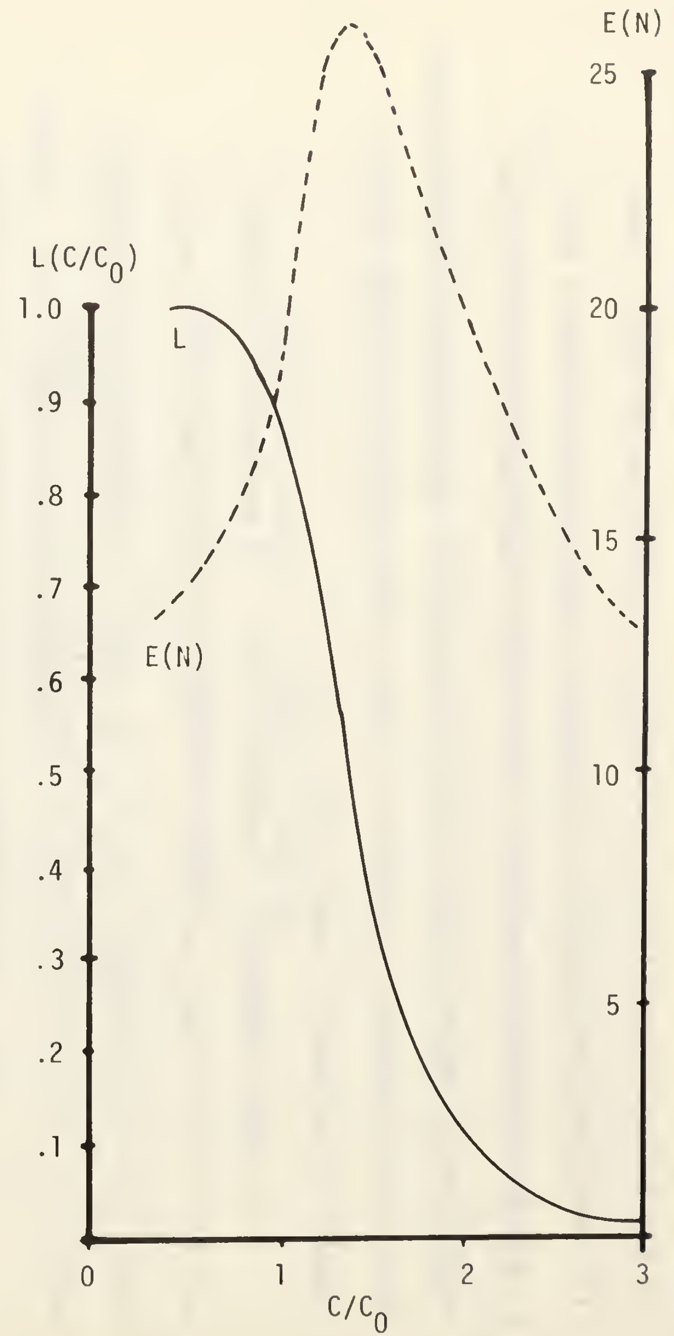

Figure 4. Approximate Characteristics of Example Sequential Proportion Test.

In this figure we show a hypothetical sequence of values $(n, \sum_{i=1}^{n} X_i)$ which led to acceptance of H_0 at the tenth observation. The O.C. curve and expected sample size curve are shown in figure 4, as functions of C/C_0. These curves are based on Wald approximations.

It is interesting to compare the O.C. curves and expected sample size ($E(N)$) curves for the Rayleigh and proportion tests, for the example case. Note that the O.C. curves are quite similar, but that the $E(N)$ curve for the proportion test is roughly twice as high as that for the Rayleigh test. For example under H_0 the sequential proportion test has $E_0(N) \approx 18$ (a nonsequential proportion test requires $n = 97$), whereas the Rayleigh has $E_0(N) = 8.8$; under the alternative hypothesis the proportion test has $E_a(N) \approx 20$ whereas the Rayleigh has $E_a(N) = 10$; the maximal expected sample size for the proportion test is approximately 26 at $C = 1.4C_0$ whereas it is 15 at $C = 1.5C_0$ for the Rayleigh. One might interpret these comparisons as the "cost" of using the "nonparametric" procedure if the parametric procedure were appropriate. Since these relationships are probably typical, we may assert that the "nonparametric" procedure gives protection against invalid testing due to departures from the parametric assumption, at the expense of additional observations (on the average).

As is typically true of tests of this type, the expected sample sizes will generally decrease with any of the following:

a. increase in α

b. increase in β

c. increase in the ratio C_1/C_0*.

---

*A decrease in $E(N)$ for true C values not near the mid point of (C_0, C_1).

IV. CONCLUSIONS.

The tests discussed above should prove useful in testing whether CEP specifications are being met under various test conditions. The "nonparametric" test, the sequential proportion test, requires more observations (on the average) than its parametric competitor, but is not invalidated by departures from the parametric assumptions required by the latter. Indeed, with the sequential proportion test, one does not have to be concerned with most of the "model validation" problems usually encountered in performing a statistical analysis. (All that is required is that the observations be independent, and that the test conditions be reasonably homogeneous.

Research is currently being conducted on a third sequential test which may prove useful in the context considered here. This test will be based on sequentially observed sample medians, with the null hypothesis being rejected when the sample median falls too far from $\sqrt{C_0}$. A second area of research effort is concerned with methodology for combining results observed under the various test conditions into an overall test of performance. Results of these efforts will be reported in a future paper.

REFERENCES

[1] D. R. Barr and M. F. Jordan, "A Power Distribution Model for Weapons Accuracy," Naval Postgraduate School Technical Report NPS-55BN74011 (1974).

[2] A. Wald, Sequential Analysis, John Wiley and Sons, New York (1947).

[3] M. W. Gavlak, "An Examination of the Wald Stopping Bounds for the Sequential Probability Ratio Test," Naval Postgraduate School Thesis (1970).

[4] E. L. Lehmann, Testing Statistical Hypotheses, John Wiley and Sons, New York, (1959).

DISTRIBUTION LIST

Defense Documentation Center Cameron Station Alexandria, Virginia 22314	12
Office of Naval Research Department of the Navy Arlington, VA 22217	2
Library (Code 0212) Naval Postgraduate School Monterey, CA 93940	2
Dean of Research Code 023 Naval Postgraduate School Monterey, California 93940	1
Library (Code 55) Department OR/AS Naval Postgraduate School Monterey, California 93940	2
William F. Trisler NAVELEX Code 05431 Washington, D. C. 20306	1
Major Lou Simpleman USMC MCTSSA, Sub Unit One Marine Corps Air Facility (Helicopter) Santa Ana, California 92709	2
Captain Ted Bean USMC MCTSSA, Sub Unit One Marine Corps Air Facility (Helicopter) Santa Ana, California 92709	1
Professor T. D. Burnett Code 55Za Naval Postgraduate School Monterey, California 93940	1
Professor D. R. Barr Code 55Bn Naval Postgraduate School Monterey, California 93940	10